Differential game theory for versatile physical human-robot interaction

Article (Accepted Version)

Differential game theory for versatile physical human-robot interaction

Y Li[1,2], G Carboni[1], F Gonzalez[1], D Campolo[3] and E Burdet[1,3]

1: Department of Bioengineering; Imperial College of Science, Technology and Medicine; SW7 2AZ London, UK

2: Department of Engineering and Design; University of Sussex; BN1 9RH Brighton, UK

3: School of Mechanical and Aerospace Engineering; Nanyang Technological University; Singapore 639798

Abstract

The last decades have seen a surge of robots working in contact with humans. However, until now these contact robots have made little use of the opportunities offered by physical interaction and lack a systematic methodology to produce versatile behaviors. Here we develop the first interactive robot controller able to understand the control strategy of the human user and react optimally to their movements. We demonstrate that combining an observer with a differential game theory controller can: induce a stable interaction between the two partners; precisely identify each other's control law; and allow them to successfully perform the task with minimum effort. Simulations and experiments with human subjects demonstrate these properties and illustrate how the new controller can induce different representative interaction strategies.

Introduction

Traditional robotic manipulators, such as those used in the automotive industry, are potentially dangerous for human workers and thus are kept separated from them. However, robotic systems have been increasingly used in recent decades to work in physical contact with humans. While the potentialities of such *contact robots* are suggested by the benefits observed during physical tasks carried between humans [1, 2], they have so far made little use of the opportunities of *interactive control* [3]: either they are fully controlled by the operator (as in master-slave behavior mode [4, 5, 6]) or, conversely, their control law does not adapt during movement to the interaction with the user, e.g. in typical rehabilitation robots [7]. To efficiently assist a human user, a contact robot should arguably react and adapt to their specific motion behavior according to a desired *control strategy*.

How should a contact robot be controlled to provide a stable and appropriate response to a user with unknown dynamics during various activities ranging from sport training, to physical rehabilitation and shared driving [8, 9, 10]? Specific human-robot interactions have been studied [11, 12] but a general framework for interactive control is still missing. It has been suggested that *differential game theory* (GT) can be used as a framework to describe various interactive behaviors between a robot and its human user [13]. However, GT typically assumes knowledge of the partner's dynamics and control strategies [14, 15], while a contact robot cannot a-priori know the sensorimotor control of the human user.

On the other hand, the study of physical tasks between humans has revealed how the capability to understand the partner's sensorimotor control and to adapt one's own control is key to interaction benefits [16]. Some recent papers have studied how finite games can deal with limited knowledge of the partner [17, 18], but it is unclear how the available techniques should be transferred to the continuous control of the interaction with a robot. A method to model the interaction force of a contact robot with its user has been introduced in [19], but in an approach requiring force sensing and restricted to the *collaboration* between equals [13]. However, different control strategies are required for various tasks, such as *less-than-needed assistance* to promote successful learning in physical rehabilitation [20], or a deceptive behavior needed for *competition* that can be used to challenge a user [21].

Therefore, our goal is to develop a versatile interactive motion behavior for a robot in physical contact with a human user, with which these two agents can optimally react to each other by learning each other's control. In this purpose, we model the robot's and human's task objectives through respective cost functions in a GT framework, which enables us to specify

the desired *control strategy* [13]. However, the mechanical connection between the partners means that their individual behaviors will interfere and may lead to instability. It is unclear how to ensure stability of the human-robot system and task completion while identifying the human user's control strategy and adapting the robot's control. To address these issues for simultaneous co-adaptation and control, we propose using adaptive control [22] to estimate the partner's cost function and GT to compute a Nash equilibrium and resolve possibly conflicting situations where each partner optimises their own behavior based on the available knowledge [23]. This adaptive GT controller is validated on reaching arm movements. It does not require measurement of the interaction force, and is proved to yield a stable interaction in typical conditions. The partner's controller model can be used to specify different types of behaviors, which we illustrate in experiments with human subjects exhibiting less-than-needed assistance required for physical rehabilitation and a competitive behavior.

A game theory controller that understands the partner's control law

As human motion planning is carried out essentially in task space [24], we describe the control of a contact robot and its human user in the corresponding space $\{x\}$ of the robot's end effector, where $x \in \mathbb{R}^n$. The control can then be transferred to the joint space as described in the Methods section. The variables used throughout the paper are listed in Table 1. Let

$$v + u_h = \Psi(x, \dot{x}, \ddot{x}) \tag{1}$$

describe how a contact robot (with command $v \in \mathbb{R}^n$) and a human user (applying a force $u_h \in \mathbb{R}^n$) move the robot's dynamics $\Psi(x, \dot{x}, \ddot{x})$. Selecting the mechanical impedance process

$$v \equiv u + \Psi - I\ddot{x} - D\dot{x} \quad I, D \in \mathbb{R}^{n\times n} \tag{2}$$

to track a common and fixed target $x_d \in \mathbb{R}^n$ (with $\dot{x}_d = 0$), the control equation yields:

$$\dot{\xi} = A\xi + B(u + u_h)\,, \quad \xi \equiv \begin{bmatrix} x - x_d \\ \dot{x} \end{bmatrix}, \quad A \equiv \begin{bmatrix} 0_n & 1_n \\ 0_n & -I^{-1}D \end{bmatrix}, \quad B \equiv \begin{bmatrix} 0_n \\ I^{-1} \end{bmatrix} \tag{3}$$

where $u \in \mathbb{R}^n$ will be computed according to linear GT as described below, 0_n is the $n \times n$ zero matrix, 1_n the $n \times n$ identity matrix (with 1 as diagonal coefficients and 0 elsewhere), and I and D are the inertia and viscosity matrices specifying the desired interaction dynamics.

Table 1: Variable definitions

x	robot's end effector position
$\Psi(x, \dot{x}, \ddot{x})$	robot dynamics
v	total robot motor command
u, u_h	robot and human motor commands
I, D	desired inertia and viscosity matrices
x_d	target position
ξ	system state (with position error and velocity)
A, B	state and input matrices
$U, \widehat{U}_h$	robot and human cost functionals
$Q, \widehat{Q}_h$	robot and human state weights
$\hat{u}_h$	robot estimate of human's motor command
$L, \widehat{L}_h$	robot and human feedback gains
$P, \widehat{P}_h$	robot and human Riccati equation solutions
A_r, A_h	robot and human state matrices
$\hat{\xi}, \tilde{\xi}$	estimate of system state and state estimation error
Γ, α	arbitrary positive definite matrix and a positive scalar
$\widetilde{L}_h$	human feedback gain estimation error
$\widetilde{P}_h$	estimation error of P_h
C	desired task performance matrix
τ	joint torque
H, N, G	mass matrix, Coriolis and centrifugal matrix, gravity term
J	Jacobian matrix
F	endpoint wrench
ω, ω_h	Nash equilibrium indexes

Using non-cooperative differential GT [15], we can describe the robot's interaction with the human user eq.(3) as a game between two players who minimise their respective cost functional

$$U \equiv \int_{t_0}^{\infty} \xi^T(t)\, Q\, \xi(t) + u^T(t)\, u(t)\, dt\,, \quad \widehat{U}_h \equiv \int_{t_0}^{\infty} \xi^T(t)\, \widehat{Q}_h\, \xi(t) + \hat{u}_h^T(t)\, \hat{u}_h(t)\, dt \quad (4)$$

where $(\cdot)^T$ represents the transpose operator and $\widehat{(\cdot)}$ the estimate of $(\cdot)$. These cost functionals mean that the robot knows its cost U and estimates the cost $\widehat{U}_h$ of the human user, where $\hat{u}_h$ is the estimation of their motor command. Each of the constant weight matrices Q and $\widehat{Q}_h$ can be positive semi-definite or negative semi-definite. In these equations, each player fulfils the reaching task by minimising the error to the target while using minimal metabolic cost. Since the state ξ includes two parts: the target *error* $x - x_d$ and the velocity $\dot{x}$,

Q and $\widehat{Q}_h$ include two components corresponding to position regulation and viscosity, respectively. Based on differential GT for linear systems [25], the following control inputs of the robot and its modelled human user minimise the cost functionals eq.(4) in the sense of the Nash equilibrium:

$$u = -L\,\xi\,, \quad L \equiv B^T P\,, \tag{5a}$$

$$A_r^T P + P A_r + Q - P B B^T P = 0_{2n}\,, \quad A_r \equiv A - B\widehat{L}_h\,, \tag{5b}$$

$$\hat{u}_h = -\widehat{L}_h\,\xi, \quad \widehat{L}_h \equiv B^T \widehat{P}_h\,, \tag{5c}$$

$$A_h^T \widehat{P}_h + \widehat{P}_h A_h + \widehat{Q}_h - \widehat{P}_h B B^T \widehat{P}_h = 0_{2n}\,, \quad A_h \equiv A - BL\,, \tag{5d}$$

where $L \equiv [L_e\,, L_v]$ (obtained using the solution of the Riccati equation eq.(5b) and eq.(5a)) is the feedback gain of the robot's control and $\widehat{L}_h \equiv [L_{h,e}\,, L_{h,v}]$ (using eqs.(5d, 5c)) is the robot's estimate of the human feedback gain. Note how the robot's and modelled human's control depend on each other through A_r and A_h, characterising the coupled optimisation.

Importantly, the human user's control gain L_h is not a-priori known to the robot. However it can be estimated from eq.(3) without requiring force sensing (i.e. u_h) for actions along non-gravity affected directions, while ensuring stability of the closed loop system. In this purpose we first express how the estimation $\hat{u}_h$ affects the state ξ to $\hat{\xi}$:

$$\dot{\hat{\xi}} = A\hat{\xi} + B(u + \hat{u}_h) - \Gamma\tilde{\xi}\,, \quad \tilde{\xi} \equiv \hat{\xi} - \xi\,. \tag{6}$$

We use this equation as an observer for u_h, where Γ is a matrix to make $\Gamma - A$ positive definite, $\hat{\xi}$ the estimate of ξ and $\tilde{\xi}$ the state estimation error. Subtracting eq.(3) from eq.(6) then yields

$$\dot{\tilde{\xi}} = (A - \Gamma)\tilde{\xi} - B\widetilde{L}_h\xi\,, \quad \widetilde{L}_h \equiv \widehat{L}_h - L_h\,. \tag{7}$$

As the state estimation error $\tilde{\xi}$ is due to the error $\widetilde{L}_h$, we develop an update law for $\widehat{P}_h$ (thus for $\widehat{L}_h \equiv B^T\widehat{P}_h$) to minimize $\tilde{\xi}$. In this purpose, we consider the Lyapunov function candidate

$$V \equiv \frac{1}{2}\,\xi^T\xi + \frac{1}{2}\,\tilde{\xi}^T\tilde{\xi} + \frac{1}{2\alpha}\,\mathrm{tr}\Big(\widetilde{L}_h^T\widetilde{L}_h\Big)\,, \quad \alpha > 0\,, \tag{8}$$

where $\mathrm{tr}(M)$ is the trace of matrix M. Its time derivative yields

$$\begin{aligned}
\dot{V} = \;& \xi^T\dot{\xi} + \tilde{\xi}^T\dot{\tilde{\xi}} + \frac{1}{\alpha}\,\mathrm{tr}\Big(\widetilde{L}_h^T\dot{\widetilde{L}}_h\Big) \\
= \;& -\xi^T(BL + BL_h - A)\xi + \tilde{\xi}^T\Big[(A-\Gamma)\tilde{\xi} - B\widetilde{L}_h\,\xi\Big] + \frac{1}{\alpha}\,\mathrm{tr}\Big(\widetilde{L}_h^T\dot{\widetilde{L}}_h\Big) \\
\equiv \;& -\xi^T(BL + BL_h - A)\xi - \tilde{\xi}^T\,\Gamma_0\,\tilde{\xi} - \tilde{\xi}^T B\widetilde{L}_h\,\xi + \frac{1}{\alpha}\,\mathrm{tr}\Big(\widetilde{L}_h^T B^T\dot{\widetilde{P}}_h\Big) \\
= \;& -\xi^T(BL + B\widehat{L}_h - A)\xi - \tilde{\xi}^T\,\Gamma_0\,\tilde{\xi} - (-\xi + \tilde{\xi})^T B\widetilde{L}_h\,\xi + \frac{1}{\alpha}\,\mathrm{tr}\Big(\widetilde{L}_h^T B^T\dot{\widetilde{P}}_h\Big)
\end{aligned} \tag{9}$$

for some positive definite matrix $\Gamma_0 \equiv \Gamma - A$, where we have used the system dynamics eq.(3) and estimation error dynamics eq.(7). Setting

$$\dot{\widehat{P}}_h \equiv \alpha\,(\tilde{\xi} - \xi)\xi^T\,, \quad \alpha > 0\,, \tag{10}$$

the last two terms of eq.(9) cancel, yielding a negative $\dot{V}$, while $\widehat{Q}_h$ can be computed from $\widehat{P}_h$ using eq.(5d). Equation (10) is critical as it enables identifying the control of the partner in order to adapt one's own control using eq.(5). The above derivation leads to the following *Theorem* (demonstrated in the Methods) which summarises our results on co-adaptation and simultaneous control:

Considering the system of eq.(3), if the robot and human estimate the partner's controller and develop their own control as in eqs.(5),(6),(10), then:

- *the closed-loop system is stable and u, u_h, ξ are bounded;*
- *the partners' controllers and cost functions converge to the correct values if ξ is persistently exciting;*
- *the Nash equilibrium is achieved, i.e., the cost functions in eq.(4) will increase if either of the control inputs differs from those in eq.(5).*

Importantly, the interactive controller with co-adaptation and partner's identification described in eqs.(5),(6),(10) can be used to implement representative interaction control behaviors such as those identified in [13]. *Coactivity*, when two agents ignore each other's dynamics, is implemented using for their control the respective part of eqs.(4),(5) with $\widehat{P}_h \equiv 0$ and $\widehat{P} \equiv 0$ where $\widehat{P}$ is human's estimate of P, yielding two independent linear quadratic regulators (LQR). *Collaboration* between two partners without hierarchy arises when they contribute to the dynamics eq.(3) using each controller of eq.(5) with positive definite $\{Q, Q_h\}$. *Competition* can be implemented with a negative definite gain matrix Q for the robot in order to challenge the human user achieving their goal by pushing them away from it, which will promote active learning [20]. *Cooperation* arises when the two agents take complementary roles, which can be defined through the *sharing rule*

$$Q + Q_h \equiv C \tag{11}$$

enabling the GT controller to continuously modify the contributions between the partners [26]. Different from collaboration, cooperation fixes the task performance through the total weight C, and uses eq.(11) to share the effort. A special case of cooperation, *assistance*, arises when C is set to let the robot

fulfil the task alone without interaction with a human user. As humans tend to relax during motor actions [27, 28], the robot will gradually take over the whole task load while the human will become passive. Another special case of cooperation particularly important in physical training is *assist-less-than-needed* [20, 7], designed to keep a human trainee engaged during sport practice or physical rehabilitation. This interaction control mode is implemented by setting C to make the robot short of reaching the target alone, which will bring the human to increase their effort in order to complete the reaching.

A pseudo-code summarizing the steps of the proposed algorithm is

---

Input: Current state ξ, target x_d.
Output: Robot's control input u, estimated human's cost function weight $\widehat{Q}_h$ in eq.(5d).
begin
Define the target position x_d, initialize Q, $\widehat{Q}_h$, u, $\hat{u}_h$, $\hat{\xi}$, $\widehat{P}_h$, set the parameters Γ in eq.(6), α in eq.(10), and the terminal time t_f of one trial. In the case of *Cooperation* set also C.
while $t < t_f$ **do**
- Measure the position x and velocity $\dot{x}$, and form the state ξ.
- Calculate the estimation error $\tilde{\xi}$ and the estimated state $\hat{\xi}$ using eq.(6): $\tilde{\xi} \equiv \hat{\xi} - \xi$, $\dot{\hat{\xi}} = A\hat{\xi} + B(u + \hat{u}_h) - \Gamma\tilde{\xi}$.
- Update the matrix $\widehat{P}_h$ using eq.(10): $\dot{\widehat{P}}_h \equiv \alpha(\tilde{\xi} - \xi)\,\xi^T$. Then compute the estimated human's control gain $\widehat{L}_h \equiv B^T \widehat{P}_h$ and motor command $\hat{u}_h \equiv -\widehat{L}_h \xi$.
- Solve the Riccati equation in eq.(5b) to yield P, and compute the robot's control input $u = -B^T P \xi$.
- Calculate the estimated human's cost function weight $\widehat{Q}_h$ according to the Riccati equation in eq.(5d).

In the case of *Cooperation*, use eq.(11) to adapt the corresponding cost function weight $Q = C - \widehat{Q}_h$.

Simulations and experiments

To test the interactive control and co-adaptation of the two agents, we simulated a neurorehabilitation scenario of arm reaching movements back and forth between $-10cm$ and $+10cm$ (Fig.1). The robot dynamics were simulated using eqs.(5),(6),(10) to generate u, the simulated human dynamics used a similar set of equations to generate u_h, and both of these inputs were

used to drive the task dynamics according to eq.(3). Both agents were assumed to have no initial knowledge of the partner's control (thus initially $\widehat{P}_h \equiv 0$ and $\widehat{P} \equiv 0$) while eqs.(6),(10) were used to estimate it. Fig.1a shows that the collaborative GT controller yields a stable behavior fulfilling the reaching task. With this controller the human control gain was identified with an error smaller than 5% of the real value in $10s$.

Could the reaching task also be fulfilled using a more simple control law neglecting the partner's control? We addressed this question by setting $\widehat{P}_h = 0$ and $\widehat{P} = 0$ in eq.(5) for the simulated robot and human agents, respectively. This leads to the simplified solution of two agents with independent LQR ignoring the interaction with the partner. Simulation of the reaching task, displayed in the dashed lines of Fig.1a, suggests that while the task is generally fulfilled with the independent LQR, this would however require larger control gains L_h thus a larger effort compared with the GT controller. Both the LQR gains and the GT gains depend on the predefined cost functionals. To investigate the effect of the costs systematically, we examined the gains in LQR and GT when Q_h varies from $Q/10$ to $10Q$. We see in Fig.1b that the LQR gains are always larger than the GT gains, with the difference becoming smaller when Q_h increases, as the robot's relative influence decreases. The GT controller considers the interaction with the simulated human (through eq.(5b)) and computes the minimal force for the robot to fulfil the task with them. In turn the simulated human can also minimise their effort.

Is the partner's control identification really necessary, or could any approximate partner's model be used instead? To test this question we carried out reaching simulations with a biased identification. Figs.1c,d show that a bias in the partner's velocity estimation can bring instability to the system and prevent it from reaching the target. The Theorem's demonstration in the Methods shows that both the bias level leading to instability and the relative effect of position and velocity gains on stability depend on the system matrix A and on the cost functionals. Using these cost functionals with the values set in the Methods, both position and velocity gains are positive regardless of the estimation bias. However the system matrix A is unstable due to damping. Therefore the closed-loop system will become unstable if the velocity gain cannot compensate for the damping due to estimation of velocity gain with a bias larger than $50N/m$. Figs.1c,d illustrate the system's behavior in the stable case of a $30N/m$ bias and in the unstable case of a $50N/m$ bias, respectively. These results illustrate the importance of accurately identifying the partner's control based on a sound analysis of the dynamics involved, as was developed in eqs.(6-10).

In order to test the ability of the controller to adapt to any partner's motor condition, we simulated the scenario of an individual with motor weakness who uses the robot to help recover their motor function (Fig.2a,b,c). We assumed that the simulated human improves their arm control by increasing the weight Q_h trial after trial as described in the Methods. Robot interaction was adapted after each trial based on the estimated weight of the simulated human's cost function. To promote active learning while still assisting a weak user, the simulated robot used an *assist-less-than-needed cooperation* strategy, with eq.(11) and C making the robot short of reaching the target alone. We see that while the reaching task is completed with similar performance in all trials (Fig.2c), this is achieved with different contributions of the simulated human and robot (Figs.2a,b). Specifically, Figs.2a,b show how the robot updates its estimate of the simulated human motor control weight Q_h and reduces its own weight Q correspondingly.

Interestingly, it appears in Fig.2b that the robot's weight in the fifth iteration decreases to a negative value, i.e. the robot provides resistance (instead of assistance) to the simulated human. This form of *competition* is desirable as it will let the user try to improve performance and be engaged in the physical training. However, the robot will not force the trainee beyond their motor capabilities, as it will keep compensating for the missing dynamics through eq.(11). In turn, this illustrates the capability of the adaptive GT controller to induce various types of interactive behaviors, namely here *less-than-needed assistance* and *competition*. The simulations results also illustrate that the two components in the weight matrices for reaching and viscosity (that may correspond to spasticity in a neurologically impaired user [29]) can be modulated separately, which makes it possible to implement requirements in various conditions.

Considering the scenario of an impaired user such as a stroke survivor using the robot to train their motor control, their behavior may be erratic with irregular progresses and setbacks over consecutive training sessions. To test the capability of our interactive controller to deal with such erratic behaviors, we simulated a human user with performance varying trial after trial, which is implemented by random changes in the corresponding cost function. Fig.2d shows that even in this situation the robot can adapt to the changing performance by updating its own cost function weight in a suitable way trial after trial.

How will our designed interaction behavior and adaptation work with real human users? We tested this by implementing less-than-needed assistance control during arm reaching (on the robotic interface of Fig.3a) as illustrated in Fig.3b. The red traces in Fig.3c illustrate the robot's reaching behavior without interaction with a user. We can observe a nearly constant motion

in consecutive trials with estimated weight of the human user at ≈ 0, as is expected. The movement is short of the $3cm$ target due to the low C parameter set in the controller via eq.(11). The pink traces exhibit the adaptation taking place in the *healthy subject scenario* where the robot assists a representative user instructed to complete the reaching task as best as possible. We observe that the reaching task is effectively carried out, while the robot successfully identifies the increasing contribution of the human user and lowers its assistance correspondingly. Interestingly, the jumps in the weights at the intervals when the movement direction changes stimulate the convergence of the estimated values to the real ones. When the subject is instructed to keep the arm passive, in a *weak arm scenario*, the robot observes this behavior with a low estimated weight Q_h and increases its motor weight Q correspondingly.

To test whether the behavior with our GT controller depends on subject specific parameters, we repeated the same experiment with 10 healthy subjects, who carried out five trials with 3 back and forth reaching movements, in each of the *healthy* and *weak arm* conditions. Fig.3d shows that the controller does not prevent natural motion variability in consecutive trials, which may facilitate learning [24]. For each subject the $Q_{h,e}$ value converged to reliable values, yielding values with *healthy* > *no interaction* (t-test, $\max\{p_i\} < 0.009, i = 1 \dots 10$) and with *weak arm* < *no interaction* (t-test, $\max\{p_i\} < 0.012, i = 1 \dots 10$). This demonstrates that our algorithm is able to clearly identify the different behaviors of each subject and adapt the GT control correspondingly. This is further illustrated in Fig.3e, where we asked a representative subject to frequently vary his behavior, as can be expected in neurologically impaired individuals. The result shows that the controller is able to catch these behavioral changes in a suitable way. Altogether, these results demonstrate the efficiency of the simple regulation of assistance during cooperation of the human and robot through the sharing rule eq.(11) with the GT controller of eqs.(5),(6),(10).

Discussion

Contact robots that physically interact with humans are being increasingly used in industry and for physical training, but they lack a systematic methodology to produce versatile interactive behaviors. Typical rehabilitation robots to help limbs movement training [7], or intelligent industrial systems to support heavy objects against gravity and facilitate their manipulation [30], use a controller independent of the human user. This co-activity type of interaction strategy, which does not require the robot to observe the human user behavior, works only as long as the robot's task corresponds to the user's

task. Conversely, in master-slave control the robot fully follows the user's movement. Master-slave control has been used for teleoperated robots, in force extenders exoskeletons [31], for redirecting movements to admissible areas in Northwestern University's cobots [32], or for robot-assisted mirror physical therapy where the unaffected arm of an hemiplegic patient drives the impaired arm [4, 33]. These two simple interaction behaviors, where the robot either ignores the human operator control (in co-activity) or blindly follows it (in master-slave control) have led to successful systems for specific tasks. However, the example of sophisticated interactions between humans promise a more versatile and flexible interaction optimally combining the two partners' capabilities [2, 1, 16]. Such strategies require an efficient process to understand the partner's control, which has not been developed in previous interactive controllers.

To design such versatile interactive control, this work thus addressed the basic human-robot interaction control issue, where a pair of physically connected agents have to understand each other's control and update their own control in order to successfully complete a reaching task together. This simultaneous partner's control identification and co-adaptation has not been addressed in previous works [19, 34], and it was unclear whether this could lead to an equilibrium or would result in instability. Integrating differential GT and an observer, the adaptive controller presented in this paper provides a stable solution to this problem, with bounded control signals, identification of the partner's control law and minimisation of the individual cost in both agents in the sense of the Nash equilibrium. The simulation results showed that the proposed method enables the robot and human to estimate each other's cost function during interaction accurately and without force sensing, as well as to adapt their control correspondingly to fulfil the common task while guaranteeing a specific interactive behavior. The adaptive properties of the control and specific behaviors were further shown in a robotic implementation with human users mimicking physical training for motor recovery.

Does the success of this implementation require that the humans' motor control corresponds to GT? It is still unknown whether the human central nervous system behaves as predicted by GT to physically interact with other humans, although there is evidence that it considers the partner's sensorimotor control during common task performance [16] and may carry out behavioral decisions that conform to GT [35]. However, irrespective on whether humans' sensorimotor control corresponds to GT, the controller of eqs.(5),(6),(10) will induce both a stable interaction with the human user (as it adapts its control gain to compensate for the human bounded but possibly unstable control gain), and an interactive behavior corresponding to

a desired control strategy. In turn, the algorithm introduced in this paper may be used to infer how human motor control corresponds to GT, and to determine which characteristic behaviors are adopted during various tasks.

The simulation and implementation results demonstrate that the new adaptive GT controller can flexibly implement different representative behaviors of physical interaction between two agents as proposed in [13, 1]. These include the collaboration or competition between partners without a hierarchy, the cooperation of complementary partners, as well as assist-less-than-needed assistance promoting engagement and active training. Therefore the present work achieved the programme of [13], by explicating the critical partner's *action understanding* necessary for an efficient interactive control and by developing the versatile differential GT control algorithm to implement these behaviors. In contrast the GT controller of [19] could only implement a collaboration and required force sensing.

The algorithm for the simultaneous partner's control identification and co-adaptation, which was developed for a reaching movement in this paper, may be generalised to more complex movements such as tracking [16]. It can be used to regulate the interaction between any two autonomous agents, such as two robots, or to design human-robot interaction as was demonstrated in this paper. It may further serve as a tool to model bimanual control with brain interhemispheral communication in both healthy individuals and neurologically impaired (e.g. cerebral palsy) subjects, as well as to model the control of physical interaction in a human pair.

The experiments of this paper illustrated that the presented algorithm can also be directly used as a dynamic environment for training impaired reaching behavior e.g. after a stroke. Interesting properties of the novel adaptive controller for training arm movements include the dynamic interaction not preventing natural motion variability in consecutive trials, as well as the stable and less-than-needed motion assistance, from nearly complete motion guidance to competition; both of these factors are critical for successful neurorehabilitation [36, 37]. The experiment demonstrated the adaptability of the robot control yielding a challenging but supportive training environment and inducing active inconspicuous adaptation.

Methods

Joint space control

The rigid body dynamics of many serial mechanisms can be represented through an equation of the form [38]

$$\tau = H(q)\,\ddot{q} + N(q,\dot{q})\,\dot{q} + G(q), \tag{12}$$

where q is the joint space vector, H the mass matrix, $N\dot{q}$ the Coriolis and centrifugal forces (which can be completed with other velocity dependent forces such as damping), and G the gravity term. This equation can be transformed to a position dependent force $\Psi(x, \dot{x}, \ddot{x})$ using kinematic transformations

$$\dot{x} \equiv J(q)\,\dot{q}\,, \quad x = \int \dot{x}\,\mathrm{d}t\,, \quad \ddot{x} = \dot{J}\dot{q} + J\ddot{q}\,, \quad \tau = J^T F\,, \tag{13}$$

where x is the robot end effector pose and F the endpoint wrench. For a redundant mechanism, the pseudo-inverse $J^\dagger = J^T(JJ^T)^{-1}$ of the Jacobian J can replace J^{-1} to compute the robot endpoint force, as well as its joints angle, angular velocity and acceleration. In the case of a parallel mechanism, the dynamics will generally already be in a form $\Psi(x, \dot{x}, \ddot{x})$ [39].

Theorem demonstration

Substituting update law eq.(10) into eq.(9) yields

$$\dot{V} = -\xi^T(BL + B\widehat{L}_h - A)\xi - \tilde{\xi}^T\,\Gamma_0\,\tilde{\xi}\,. \tag{14}$$

$BL + B\widehat{L}_h - A$ is positive definite if Q is positive definite, according to eq.(5b). As Γ_0 is also positive definite, it follows $\lim_{t\to\infty}\|\xi(t)\| = 0$ and $\lim_{t\to\infty}\|\tilde{\xi}(t)\| = 0$. Therefore ξ is bounded and $\lim_{t\to\infty}\|\dot{\tilde{\xi}}(t)\| = 0$ if $\lim_{t\to\infty}\|\tilde{\xi}(t)\|$ exists. According to the estimation error dynamics eq.(7) we thus have $\lim_{t\to\infty}\widetilde{L}_h(t)\,\xi(t) = 0_{n\times 1}$. If ξ is *persistently exciting*, i.e. there exist positive constants t_0, T_0 and β with $\int_t^{t+T_0}\xi(s)\,\xi^T(s)\,\mathrm{d}s \geq \beta\,1_{2n}\ \forall t \geq t_0$, it follows $\lim_{t\to\infty}\|\widetilde{L}_h(t)\| = 0$, thus $\widetilde{L}_h(t)$ is bounded. L_h is assumed to be bounded since it is human's control gain. Therefore, $\widehat{L}_h$ is also bounded. According to eq.(5b), A_r is bounded so also P, L and u. The boundedness of u_h follows from the boundedness of L_h and ξ.

Using eq.(5d), we can compute the estimation error $\widetilde{Q}_h = \widehat{Q}_h - Q_h$ which is due to the errors $\widetilde{P}_h$ and $\widetilde{P}$. As both of these errors converge to zero with the human and robot estimations, we have $\widetilde{Q}_h \to 0$, i.e., the human's unknown cost function weight Q_h is identified by the robot. It can be shown similarly that the robot's cost function weight Q is identified by the human.

Multiplying the robot equation in eq.(5b) by $\hat{\xi}^T$ on the left side and by $\hat{\xi}$ on the right side, and considering the estimation error dynamics eq.(7) yields

$$0 = \hat{\xi}^T Q\hat{\xi} + \hat{\xi}^T PBB^T P\hat{\xi} + 2\hat{\xi}^T P(\dot{\hat{\xi}} + \Gamma\tilde{\xi}) \equiv \widehat{\omega}\,. \tag{15}$$

Taking the limits $\tilde{\xi} \to 0_{2n\times 1}$ and $\dot{\tilde{\xi}} \to 0_{2n\times 1}$ yields

$$\omega \equiv \xi^T Q\xi + \xi^T PBB^T P\xi + 2\xi^T P\dot{\xi} \ \to\ 0\,. \tag{16}$$

It can be shown similarly that

$$\omega_h \equiv \xi^T Q_h \xi + \xi^T P_h B B^T P_h \xi + 2\xi^T P_h \dot{\xi} \;\; \rightarrow \;\; 0 \tag{17}$$

as $\tilde{\xi}_h \rightarrow 0_{2n\times 1}$ and $\dot{\tilde{\xi}}_h \rightarrow 0_{2n\times 1}$, where $\tilde{\xi}_h = \hat{\xi}_h - \xi$ is the state estimation error of the human. $\omega_h \rightarrow 0$ and $\omega \rightarrow 0$ indicate that the Nash equilibrium is achieved for the human-robot interaction system [25].

Simulations

10 back and forth reaching movements were simulated between $-10cm$ and $+10cm$ with a total time of $40s$. The simulation parameters were set as: inertia $I = 6kg$, damping $D = -0.2N/m$, observer gains $\Gamma \equiv \text{diag}(10, 1)$, learning rate $\alpha \equiv 10^4$. The collaborative interaction of Fig.1 was controlled using $Q = Q_h \equiv \text{diag}(100, 0)$. Biased identification was simulated by adding $[0, 30]Ns/m$ and $[0, 50]Ns/m$ offsets to the estimated robot and human velocity gains, respectively.

Motor recovery corresponding to the simulation shown in Figs.2a,b,c was implemented as an iterative increase of the weight in the human's cost function with $Q_h \equiv \text{diag}(10, 0) + i\,\text{diag}(60, 0)$ where i is the trial number, while $C \equiv \text{diag}(200, 0)$ was set for the robot to fulfil the reaching task alone. The human's erratic behavior in Fig.2d was simulated by using a uniformly distributed random number ρ in the weight of the human's cost function with $Q_h \equiv \text{diag}(200, 0) + \rho\,\text{diag}(-100, 0)$.

Experiment

Setup. The validation with (healthy) human subjects was carried out using the HMan, a cable-actuated 2-DOF manipulandum [40]. The version fabricated at Imperial College has a rectangular $24 \times 28\,cm^2$ workspace and visual feedback co-located with planar arm movements provided on the screen (Figs.3a,b). The control of the 2 Maxon motors was carried out at $1000Hz$ using realtime LabVIEW while the (x, y) position was computed from the motors' optical encoders.

Task. The reaching task included 3 back and forth reaching movements between $-10cm$ and $+10cm$ with a total time of $12s$, which was long enough for the robot to identify the human user's control. This experiment was approved by the ethics committee at Imperial College (No. 16IC3580). The actual HMan robot dynamics $\Psi \equiv I\ddot{x} + D\dot{x}$ were used as the desired dynamics. The following control parameters were used throughout the experiment: desired task performance weight $C \equiv 4000$, observer gain $\Gamma \equiv 30$, learning

rate $\alpha \equiv 2\times10^4$, and initial parameter $\widehat{P}_h \equiv 0$. C was tuned so that the robot without interaction comes within $3cm$ from the target.

Protocol. Ten naive subjects without known sensorimotor impairment participated in the experiment, after having given their informed consent. Each subject was instructed to sit in front of the robotic interface facing the monitor and move the handle by following "move forward" and "backwards" instructions displayed on the monitor. Subjects were allowed to practise until they become familiar with the reaching task and the robotic interface. Subsequently, each subject performed five trials (with one trial consisting of three back and forth reaching movements) in each of the two experimental conditions *healthy* and *weak arm*, in which they were instructed to reach the target and to completely relax the arm, respectively. The control condition *no interaction* was run by the robot alone for 50 times.

Data analysis. For each subject and each condition of *healthy* or *weak arm*, a t-test was applied on the last value of $Q_{h,e}$ of the five trials to determine the difference from the *no interaction* condition. The fourth trial of subject 1 had to be discarded due to a recording error as the subject started that trial too early, so there were four values in this case. For each subject, the difference was significant with $p < 1.2\%$ and we report the maximum p value over the subjects.

Code Availability The code that supports the findings of this study is available from the corresponding authors upon reasonable request.

Identifiable Images Consent to publish identifiable images of research participants was obtained.

References

[1] Sawers, A. & Ting, L.H. Perspectives on human-human sensorimotor interactions for the design of rehabilitation robots. *J. NeuroEng. Rehabil.* **11**(1), 142 (2014).

[2] Ganesh, G., Takagi, A., Osu, R., Yoshioka, T., Kawato, M. & Burdet, E. Two is better than one: Physical interactions improve motor performance in humans. *Sci. Rep.* **4**, 3824 (2014).

[3] Jarrassé, N., Sanguineti, V. & Burdet, E. Slaves no longer: review on role assignment for human-robot joint motor action. *Adapt. Behav.* **22**(1), 70-82 (2014).

[4] Hesse, S. et al. Computerized arm training improves the motor control of the severely affected arm after stroke: A single-blinded randomized trial in two centers. *Stroke* **36**(9), 1960-1966 (2005).

[5] Hokayem, P.F. & Spong, M.W. Bilateral teleoperation: An historical survey. *Automatica* **42**(12), 2035-2057 (2006).

[6] Passenberg, C., Peer, A. & Buss, M. A survey of environment-, operator-, and task-adapted controllers for teleoperation systems, *Mechatronics* **20**(7), 787-801 (2010).

[7] Colombo, R. & Sanguineti, V. Assistive controllers and modalities for robot-aided neurorehabilitation. in *Rehabilitation Robotics* (Colombo, R & Sanguineti, V. ed.), 63-74 (Elsevier, 2018).

[8] Marchal-Crespo, L. et al. The effect of haptic guidance and visual feedback on learning a complex tennis task. *Exp. Brain Res.* **231**(3), 277-291 (2013).

[9] Díaz, I., Gil, J.J. & Sánchez, E. Lower-limb robotic rehabilitation: Literature review and challenges. *J. Robotics* **2011**, 1-11 (2011).

[10] Na, X. & Cole, D.J. Linear quadratic game and noncooperative predictive methods for potential application to modelling driver-AFS interactive steering control. *Vehicle Sys. Dyn.* **51**(2), 165-198 (2013).

[11] Music, S. & Hirche, S. Control sharing in human-robot team interaction. *Annu. Rev. Control* **44**, 342-354 (2017).

[12] Khoramshahi, M. & Billard, A. A dynamical system approach to task-adaptation in physical human-robot interaction. *Auton. Robot.*, 1-20 (2018).

[13] Jarrassé, N., Charalambous, T. & Burdet, E. A framework to describe, analyze and generate interactive motor behaviors. *PLoS ONE* **s7**(11), e49945 (2012).

[14] Starr, A.W. & Ho, Y.-C. Nonzero-sum differential games. *J. Optimiz. Theory App.* **3**(3), 184-478 (1969).

[15] Basar, T. & Olsder, G.J. *Dynamic noncooperative game theory, 2nd edition* (Society for Industrial and Applied Mathematics, 1999).

[16] Takagi, A., Ganesh, G., Yoshioka, T., Kawato, M. & Burdet, E. Physically interacting individuals estimate their partner's movement goal to enhance motor abilities. *Nat. Hum. Behav.* **1**, 54 (2017).

[17] Kiumarsi, B. et al. Optimal and autonomous control using reinforcement learning: A Survey. *IEEE T. Neur. Net. Lear.* **29**(6), 2042-2062 (2018).

[18] Marden, J.R., Arslan, G. & Shamma, J.S. Joint strategy fictitious play with inertia for potential games. *IEEE T. Automat. Contr.* **54**(2), 208-220 (2009).

[19] Li, Y., Tee, K.P., Yan, R., Chan, W.L. & Wu, Y. A framework of human-robot coordination based on game theory and policy iteration. *IEEE T. Robot.* **32**(6), 1408-1418 (2016).

[20] Reinkensmeyer, D.J., Burdet, E., Casadio, M., Krakauer, J.W., Kwakkel, G., Lang, C.E., Swinnen, S.P., Ward, N.S. & Schweighofer, N. Computational neurorehabilitation: modeling plasticity and learning to predict recoverys. *J. NeuroEng. Rehabil.* **13**(1), 1-25 (2016).

[21] Nierhoff, T., Leibrandt, K., Lorenz, T. & Hirche, S. Robotic billiards: understanding humans in order to counter them. *IEEE T. Cybernetics* **46**(8), 1889-1899 (2016).

[22] Slotine, J.-J.E. & Li, W. *Applied nonlinear control* (Prentice-Hall, 1991).

[23] Gajic, Z. & Qureshi, M.T.J. *Lyapunov matrix equation in system stability and control* (Elsevier, 1995).

[24] Burdet, E., Franklin, D.W. & Milner, T.E. *Human Robotics: neuromechanics and motor control* (MIT Press, 2013).

[25] Engwerda, J. Algorithms for computing Nash equilibria in deterministic LQ games. *Computat. Manage. Sci.* **4**(2), 113-140 (2007).

[26] Evrard, P. & Kheddar, A. Homotopy switching model for dyad haptic interaction in physical collaborative tasks. in *Proc. IEEE Worldhaptics*, 45-50 (2009).

[27] Emken, J.L., Benitez, R., Sideris, A., Bobrow, J.E. & Reinkensmeyer, D.J. Motor adaptation as a greedy optimization of error and effort. *J. Neurophysio.* **97**(6), 3997-4006 (2007).

[28] Franklin, D.W., Burdet, E., Tee, K.P., Osu, R., Chew, C.M., Milner, T.E. & Kawato, M. CNS learns stable, accurate, and efficient movements using a simple algorithm. *J. Neurosci.* **28**(44), 11165-11173 (2008).

[29] Levin, M.F. et al. Deficits in the coordination of agonist and antagonist muscles in stroke patients: implications for normal motor control. *Brain Res.* **853**(2), 352-369 (2000).

[30] Colgate, J.E. et al., Methods and apparatus for manipulation of heavy payloads with intelligent assist devices. *USA patent*, 7185774 (2007).

[31] Zoss, A.B., Kazerooni, H. & Chu A. Biomechanical design of the Berkeley lower extremity exoskeleton (BLEEX), *IEEE-ASME T. Mech.* **11**(2), 128-138 (2006).

[32] Peshkin, M.A. et al. Cobot architecture. *IEEE T. Robotic. Autom.* **17**(4), 377-390 (2001).

[33] Burgar, C.G. et al. Robot-assisted upper-limb therapy in acute rehabilitation setting following stroke: Department of Veterans Affairs multisite clinical trial. *J. Rehabil. Res. Dev.* **48**(4), 445-458 (2011).

[34] Chackochan V.T. Development of collaborative strategies in joint action. *Ph.D. Thesis, University of Genoa, Italy*, (2018).

[35] Braun, D.A., Ortega, P.A. & Wolpert, D.M. Nash equilibria in multi-agent motor interactions. *PLoS Comput. Biol.* **5**(8), e1000468 (2009).

[36] Hogan, N. et al. Motions or muscles? Some behavioral factors underlying robotic assistance of motor recovery. *J. Rehabil. Res. Dev.* **43**(5), 605 (2006).

[37] Kahn L.E. et al. Robot-assisted movement training for the stroke-impaired arm: Does it matter what the robot does? *J. Rehabil. Res. Dev.* **43**(5), 619 (2006).

[38] Spong, M. & Vidyasagar, M. *Robot Dynamics and Control*, (Wiley, 1989).

[39] Codourey, A. & Burdet, E. A body-oriented method for finding a linear form of the dynamic equation of fully parallel robots. *Proc. IEEE Int. Conf. Robot.* **2** 1612-1618 (1997).

[40] Campolo D. et al. H-Man: A planar, H-shape cabled differential robotic manipulandum for experiments on human motor control. *J. Neurosci. Meth.* **235**, 285-297 (2014).

Acknowledgments

We thank Jonathan Eden, Thulasi Mylvaganam, Nuria Peña Perez, Quang-Cuong Pham, Keng Peng Tee and the two reviewers for their careful reading and comments on the manuscript.

Contributions Control concepts: YL, EB; Algorithm and simulation: YL; Setup FG, DC; Experiments: GC; Results analysis: YL, GC, EB; Manuscript writing: YL, EB. All authors have read and edited the manuscript, and agree with its content.

Sponsors This research was supported in part by the European Commission grant EU-H2020 COGIMON (644727), UK EPSRC MOTION grant EP/NO29003/1 and Singapore MOE Tier1 grant RG48/17.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Yanan Li (email: yl557@sussex.ac.uk) or Etienne Burdet (eburdet@ic.ac.uk).

Data Availability The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Ethical Compliance We have complied with all relevant ethical regulations.

Ethics Committee Guidelines for study procedures were provided by Imperial College London.

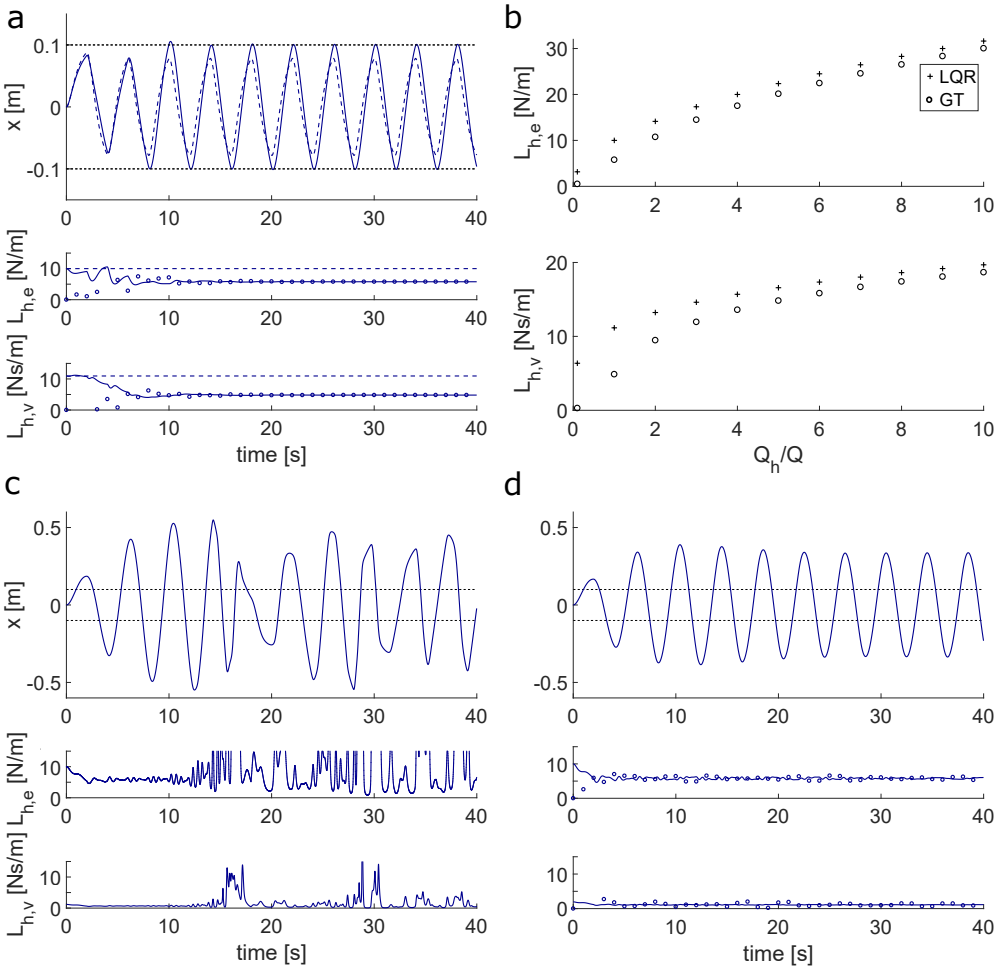

Figure 1: Simulation of an arm reaching with the game theory (GT) interactive controller. x denotes the end effector position, $L_{h,e}$ the human's position error gain and $L_{h,v}$ the velocity gain. a: The $\pm 10cm$ targets (dashed black line) can be reached with relatively small control gains with the GT controller (solid lines) since the simulated human and robot consider their partner's control and reach an equilibrium. The estimated control gains (circled lines) converge in a few seconds to the real values (i.e. the solid line). Using two independent LQR for each agent (dashed lines), the target can be almost reached but with larger control gains. b confirms that the LQR gains are larger than the GT gains for different values of Q_h relative to Q. c: The identification with a bias of $80N/m$ in the velocity gain causes control instability and prevents the controller from succeeding in the reaching task. d: The identification with a bias of $30N/m$ in the velocity gain does not cause control instability although it prevents the controller from succeeding in the reaching task. c and d illustrate the importance of using an accurate identification of the partner's control for successful performance.

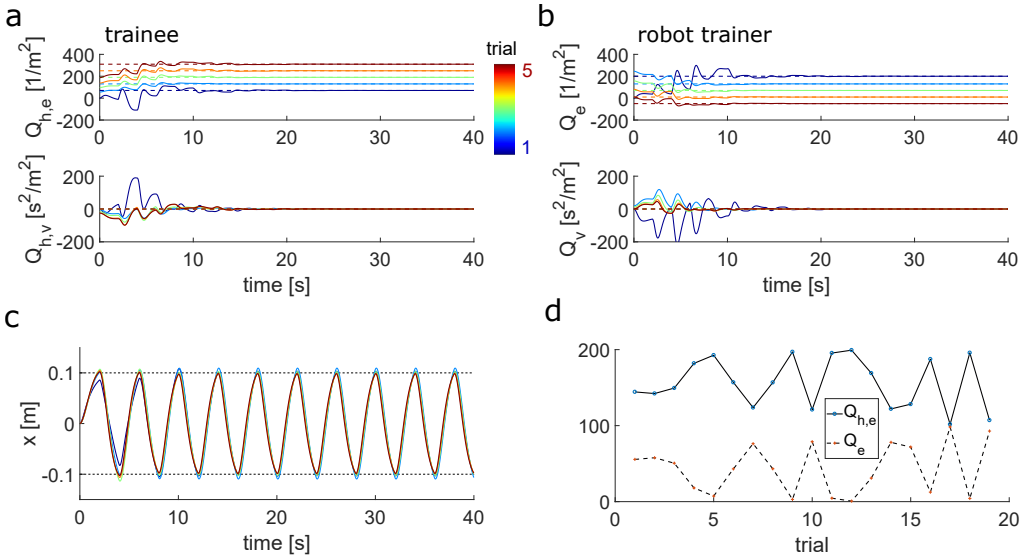

Figure 2: Simulated arm reaching training for motor recovery. a: Starting from a weak arm condition, the simulated human gradually recovers strength, which is implemented by increasing the weight in the cost function for tracking (with position weight $Q_{h,e}$ and velocity weight $Q_{h,v}$). The dashed lines correspond to real values and the solid lines to their estimation. b: The robot thus automatically reduces its assistance trial after trial by decreasing its weight in the cost function for reaching (with position weight Q_e and velocity weight Q_v). The dashed lines correspond to real values and the solid lines to their estimation. c: Position profiles under shared control of simulated human and robot remain similar in consecutive trials, despite the shift of effort towards the human trainee. d: When the simulated human's behavior is erratic (corresponding to an inconstant cost function), the robot can adapt its own cost function weight correspondingly.

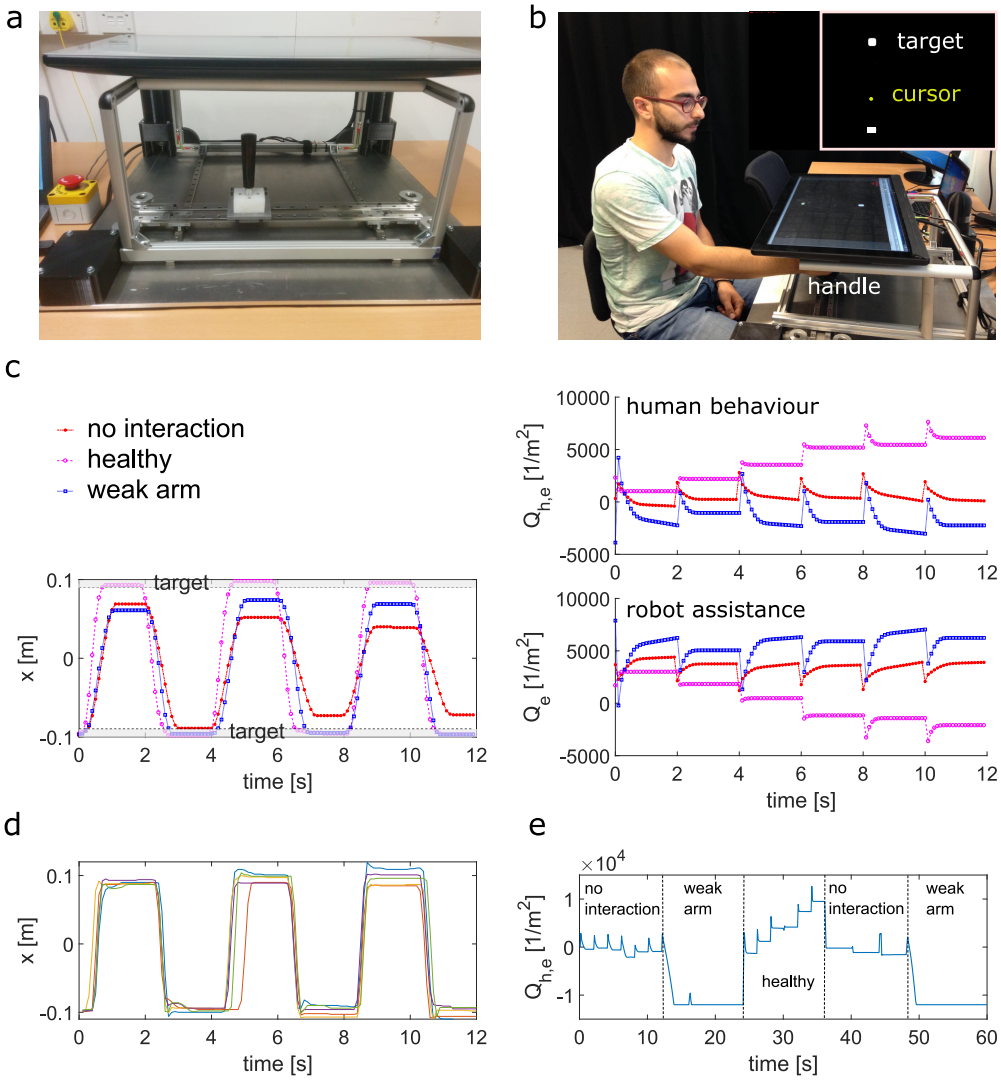

Figure 3: Adaptation of assistance to reaching experiment. a: Robotic interface to interact with horizontal arm movements used in our experiments. b: A representative human subject was instructed to adopt different roles during arm reaching movements while interacting with the robotic interface. The target and cursor are displayed on the monitor during movement. c: Movement trajectories of one representative subject in the three experimental conditions: *no interaction, healthy, weak arm*, and estimated human subject's weight and robot's weight in respective cost functions. The learning GT controller adapts *i*) the estimated human's weights in the cost functions corresponding to the user's behavior, i.e. ≈ 0 for *no interaction*, positive for *healthy* and negative for *weak arm*, and *ii)* the robot's assistance correspondingly, i.e. medium for *no interaction*, low for *healthy* and high for *weak arm*. d: One representative subject's trajectories in the five trials of the *healthy* condition. The natural motor variability in different movements is not prevented by the controller, which may facilitate learning. e: An experiment with one representative subject shows that identification of human's control weights can be achieved in various behavioral conditions.